

A SURVEY OF QUALITY MEASURES FOR GRAY SCALE IMAGE COMPRESSION

Ahmet M. Eskicioglu and Paul S. Fisher
University of North Texas
Department of Computer Science
P.O. Box 13886
Denton, TX 76203, USA
E-mail: eskiciog@ponder.csci.unt.edu
or fisher@gab.unt.edu

Abstract. Although a variety of techniques are available today for gray-scale image compression, a complete evaluation of these techniques cannot be made as there is no single reliable objective criterion for measuring the error in compressed images. The traditional subjective criteria are burdensome, and usually inaccurate or inconsistent. On the other hand, being the most common objective criterion, the mean square error (MSE) does not have a good correlation with the viewers' response. It is now understood that in order to have a reliable quality measure, a representative model of the complex human visual system is required. In this paper, we survey and give a classification of the criteria for the evaluation of monochrome image quality.

1. Introduction

There is an ever increasing demand for transmission and storage of vast amounts of information in data processing environments today. To reduce the large costs involved, data compression is a widely accepted tool which aims at minimizing the amount of data to be stored or transmitted. A variety of data compression techniques have been developed in the past few decades for different types of industrial, commercial, and educational applications. These techniques can be classified into two major categories: Lossless (exact) and lossy (inexact) [1, 2, 3]. Lossless compression is concerned with reconstructing an exact replica of the original input data stream. It is essentially used in text compression where no loss can be tolerated. Disastrous results may be encountered for even a single bit of loss in, for example, program files or database records. The techniques in this category typically reduce text size 40 to 80%, while those developed for specific applications may achieve compression over 90%. Lossy data compression causes some amount of loss which is considered to be a concession for a drastic increase in compression. Lossy compression techniques are effective and appropriate primarily for digitized voice and images for two reasons: Firstly, huge volumes of voice and images are normally generated in a typical application and, secondly, digital representation of analog signals is only an approximation, introducing a certain loss to begin with.

Numerous image compression techniques [2-6] exist today with the common goal of reducing the number of bits needed to store or to transmit images. The efficiency of a compression algorithm is generally measured using three criteria:

- 1) compression amount,
- 2) implementation complexity, and
- 3) resulting distortion.

The amount of compression can readily be obtained using several definitions, among which there are compression ratio, figure of merit, and compression percentage. Algorithmic complexity, on the other hand, can be measured by considering the data structures as well as the type and number

of operations required. The difficulty in evaluating a lossy compression algorithm comes from the fact that there is no reliable and consistent measure for determining the magnitude of distortion resulting from the loss. In other words, we lack a useful and practical measure for image quality assessment! Such a measure is not only needed for comparing images produced by different techniques, but it is also instrumental in designing image processing/compression algorithms.

In this paper, we survey the criteria available for the evaluation of monochrome image quality. In spite of the fact that some of the measures found in the literature have specifically been used for rating the performance of image processing systems, they are applicable in evaluating compression algorithms equally well.

2. Image Quality Measures

It is possible to classify image quality criteria as given in Figure 1.

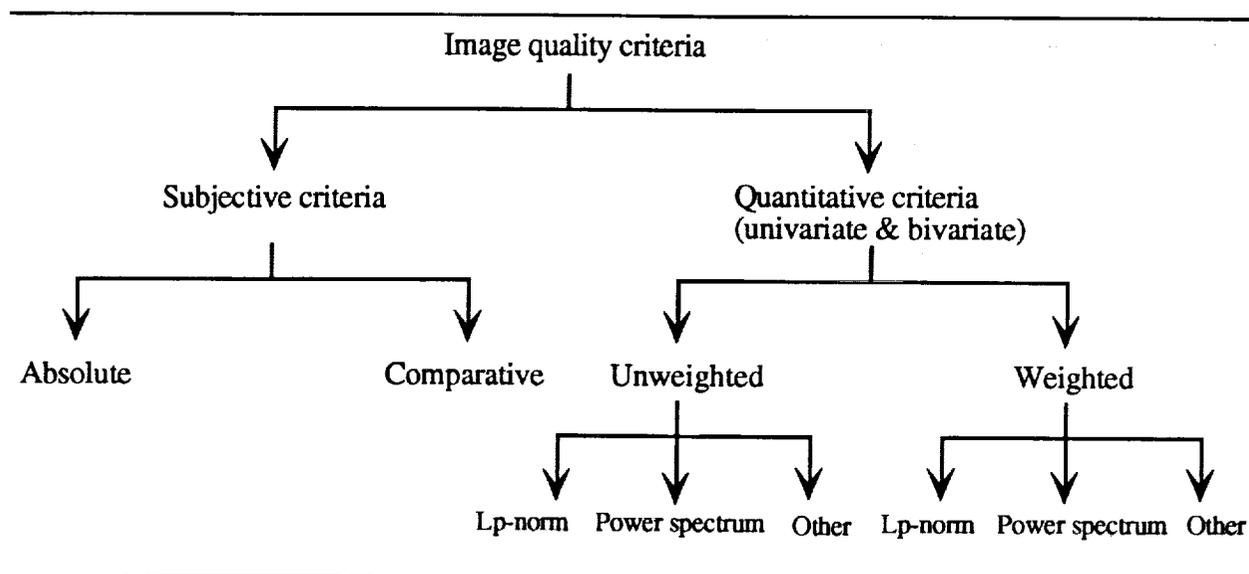


Figure 1. Classification of Image Quality Criteria

2.1 Subjective Criteria

As the final user of images are humans, the most reliable and commonly used assessment of image quality is the subjective rating by human observers. Both expert and nonexpert observers are used in experiments; nonexperts represent the average viewer while experts are believed to be able to give better, more 'refined' assessments of image quality since they have been trained and are familiar with images and their distortions.

In absolute evaluation, the observers view an image and assess its quality by assigning to it a category in a given rating scale, whereas in comparative evaluation, a set of images are ranked from best to worst by the observers. The rating scales that appear in the relevant literature [5, 12, 14, 15, 19] are listed in Table 1.

Table 1. Rating Scales Used in Subjective Evaluation

<p>A.</p> <ol style="list-style-type: none"> 5. Excellent 4. Good 3. Fair 2. Poor 1. Unsatisfactory (bad) 	<p>B.</p> <ol style="list-style-type: none"> 7. Best 6. Well above average 5. Slightly above average 4. Average 3. Slightly below average 2. Well below average 1. Worst 	<p>C.</p> <ol style="list-style-type: none"> 1. Not noticeable (perceptible) 2. Just noticeable (perceptible) 3. Definitely noticeable (perceptible) but only slight impairment 4. Impairment not objectionable 5. Somewhat objectionable 6. Definitely objectionable 7. Extremely objectionable 										
<p>D.</p> <ol style="list-style-type: none"> 3 Much better 2 Better 1 Slightly better 0 Same -1 Slightly worse -2 Worse -3 Much worse 	<p>E.</p> <ol style="list-style-type: none"> 5. Imperceptible 4. Perceptible but annoying 3. Slightly annoying 2. Annoying 1. Very annoying 	<p>F.</p> <table style="border: none;"> <tr> <td>10, 9</td> <td>Very good</td> </tr> <tr> <td>8, 7</td> <td>Good</td> </tr> <tr> <td>6, 5, 4</td> <td>Fair</td> </tr> <tr> <td>3, 2</td> <td>Bad</td> </tr> <tr> <td>1, 0</td> <td>Very bad</td> </tr> </table>	10, 9	Very good	8, 7	Good	6, 5, 4	Fair	3, 2	Bad	1, 0	Very bad
10, 9	Very good											
8, 7	Good											
6, 5, 4	Fair											
3, 2	Bad											
1, 0	Very bad											

The mean rating of a group of observers who join the evaluation is usually computed by

$$R = \left(\sum_{k=1}^n s_k n_k \right) / \left(\sum_{k=1}^n n_k \right),$$

where s_k = the score corresponding to the k th rating, n_k = the number of observers with this rating, and n = the number of grades in the scale.

Bubble sort [5, 11, 22] is another technique used in image rating. With this technique, the subject compares two images A and B from a group and determines their order. Assuming that the order is AB, he/she takes a third image and compares it with B to establish the order ABC or ACB. If the order is ACB, then another comparison is made to determine the new order. The procedure continues until all the images have been used, allowing the best pictures to bubble to the top if no ties are accepted.

It is important to note that the results of subjective rating are affected by a number of factors including

- a) type and range of images,
- b) level of expertise of the observers, and
- c) experimental conditions.

If standards can be established for these factors, the results obtained in different locations and at different times may then become comparable.

2.2 Quantitative Criteria

Quantitative measures for image quality can be divided into two classes: Univariate and bivariate [19]. A univariate measure assigns to a single image a numerical value based upon measurements of the image field, and a bivariate measure is a numerical comparison between two images.

Fidelity measurements are usually made using an array of discrete image samples, although a continuous image field can also be generated by two-dimensional interpolation of the sample array if the overhead is justified. Image error measures can be defined in either spatial or frequency domain.

Denoting the samples on the original image field as $F(j,k)$, a spatial domain, univariate quality rating may be expressed in general as

$$Q = \sum_{j=1}^M \sum_{k=1}^N O\{F(j,k)\}$$

for $N \times M$ samples, where $O\{\cdot\}$ is some operator.

Bivariate measures are more frequently used in image quality measurement. If $\hat{F}(j,k)$ denotes the samples on the degraded image field, a number of measures can be established to determine the closeness of the two image fields. The alternatives are listed below [5, 9, 12, 19, 22-25].

$$(i) \quad L_p = \left\{ (1/MN) \sum_{j=1}^M \sum_{k=1}^N |F(j,k) - \hat{F}(j,k)|^p \right\}^{1/p}$$

A major class of bivariate error measures is based on the L_p -norm. The factor p determines the relative significance of errors of different magnitudes. L_1 is the average absolute error and L_2 is the commonly used root mean square error (RMSE). As the value of p is increased, a greater relative emphasis is given to large errors in the image.

(ii) Low order moment of a power spectrum.

$$(iii) \quad K = \sum_{j=1}^M \sum_{k=1}^N F(j,k) \hat{F}(j,k)$$

This measure is obtained by discretizing the continuous cross-correlation function. It may be normalized by the reference image energy to give unity as the peak correlation:

$$NK = \frac{\sum_{j=1}^M \sum_{k=1}^N F(j,k) \hat{F}(j,k)}{\sum_{j=1}^M \sum_{k=1}^N [F(j,k)]^2}$$

(iv) Correlation quality:

$$CQ = \frac{\sum_{j=1}^M \sum_{k=1}^N F(j,k) \hat{F}(j,k)}{\sum_{j=1}^M \sum_{k=1}^N F(j,k)}$$

(v) Structural content:

$$SC = \frac{\sum_{j=1}^M \sum_{k=1}^N [F(j,k)]^2}{\sum_{j=1}^M \sum_{k=1}^N [\hat{F}(j,k)]^2}$$

(vi) Normalized absolute error between the reference and degraded image fields:

$$NAE = \frac{\sum_{j=1}^M \sum_{k=1}^N |O\{F(j,k)\} - O\{\hat{F}(j,k)\}|}{\sum_{j=1}^M \sum_{k=1}^N |O\{F(j,k)\}|}$$

(vii) Normalized mean square error:

$$NMSE = \frac{\sum_{j=1}^M \sum_{k=1}^N [O\{F(j,k)\} - O\{\hat{F}(j,k)\}]^2}{\sum_{j=1}^M \sum_{k=1}^N [O\{F(j,k)\}]^2}$$

(viii) Peak mean square error:

$$PMSE = \frac{(1/MN) \sum_{j=1}^M \sum_{k=1}^N [O\{F(j,k)\} - O\{\hat{F}(j,k)\}]^2}{A^2}$$

where A represents the maximum value of O{F(j,k)}.

The definitions used for the operator O{·} in (vii) and (viii) are

- (a) F(j,k)
- (b) [F(j,k)]^v (Power law)

- (c) $k_1 \log_b [k_2 + k_3 F(j,k)]$ (Logarithmic)
- (d) $[F(x,y) \otimes H(x,y)] \delta (x-j\Delta x, y-k\Delta y)$ (Convolution)

(ix) Laplacian mean square error:

$$\text{LMSE} = \frac{\sum_{j=1}^{M-1} \sum_{k=2}^{N-1} [O\{F(j,k)\} - O\{\hat{F}(j,k)\}]^2}{\sum_{j=1}^{M-1} \sum_{k=2}^{N-1} [O\{F(j,k)\}]^2}$$

where $O\{F(j,k)\} = F(j+1, k) + F(j-1, k) + F(j, k+1) + F(j, k-1) - 4F(j,k)$

In many applications, the mean square error (however it is defined) is often expressed in terms of a signal-to-noise ratio defined in decibels.

(x) Image fidelity:

$$\text{IF} = 1 - \frac{\sum_{j=1}^M \sum_{k=1}^N [F(j,k) - \hat{F}(j,k)]^2}{\sum_{j=1}^M \sum_{k=1}^N [F(j,k)]^2}$$

(xi) Difference $[j,k] = F(j,k) - \hat{F}(j,k)$

(xii) $\sum_{j=1}^M \sum_{k=1}^N \text{Difference } [j,k] / MN$

(xiii) $\text{Max}\{| \text{Difference}[j,k] | \}$

(xiv) Histogram of the compression error (constructed by plotting the number of x's versus x for all values of x found in the difference matrix).

(xv) Hosaka plots

(xvi) Sensitivity and predictive value positive curves

(xvii) Rate-distortion curves.

It is reported that image quality assessment can be improved by incorporating into the evaluation process some model of the HVS. The HVS is incorporated into the quality measure using two distinct approaches. In the first approach, the L_p norm (or one of its variants) is employed attaching a weight to the image samples either in the spatial or frequency domain. The second approach is concerned with weighting the digital image power spectrum.

In one of the earliest studies, the transformation

$$O\{\cdot\} = H_L(x,y) \otimes O_N\{\cdot\}$$

is used on both the continuous image field $F(x,y)$ and the degraded image field $\hat{F}(x,y)$ before applying the integral square error, where the impulse response $H_L(x,y)$ represents the lateral inhibition process, and the point nonlinearity $O_N\{\cdot\}$ models the response of the eye's photoreceptors [11]. In the Fourier domain H_L is defined as

$$a \left[c + \left(\frac{\omega}{\omega_0} \right)^{k_1} \right] \exp \left\{ - \left(\frac{\omega}{\omega_0} \right)^{k_2} \right\},$$

where $\omega = (\omega_1 + \omega_2)^{1/2}$, and $O\{\cdot\} = \{\cdot\}^{1/3}$ is chosen. The experiments show that $a = 2.6$, $c = 0.0192$, $\omega_0 = 1/0.114$, $k_1 = 1$ and $k_2 = 1.1$ are the suitable parameter values.

In another study [12] to find an objective measure which closely mirrors the performance of the human viewer, the error measure

$$E_p = \left\{ \frac{1}{m} \sum_{i=1}^m |e_i|^p \right\}^{1/p}$$

where m = number of picture elements (pels) in a picture, $e_i = x_i - \hat{x}_i$, x_i = the value of the pel in the original picture and \hat{x}_i = the value of the pel in the distorted picture, is tried for $p = 1, 2, 3, 4, 6$. The conclusion is that E_p is a very good estimate of impairment rating where the type of distortion is additive white noise. In the same study, another measure of picture impairment is obtained using

$$EM_p = \left\{ \frac{1}{m} \sum_{i=1}^m |e_i|^p / W_i \right\}^{1/p}$$

to reflect the masking effect of the signal. W_i denotes the value of the weighting function at pel i and is derived from an activity function that is a measure of the variability of the signal in the neighborhood of pel i . Three different forms of activity functions are studied:

- A_{\max} : measures the maximum signal change between any pair of pels in a neighborhood consisting of the pel being evaluated plus the eight surrounding pels.
- A_{av} : sums the deviations of the same neighborhood of points from the neighborhood average \bar{x}
- A_{df} : provides the weighted sum of the magnitude of the surrounding element difference (slope) in both the horizontal and vertical directions.

In all three cases W_i is obtained from A_i so as to span a range from 1.0 to 10.0. There is also an attempt in [12] to obtain a local measure of image quality. Relying on the postulate that the viewer rates the image by some weighted average of the worst two or three patches, Limb divides the image into a rectangular array of squares and calculates a local measure for each square with and without masking. He also tries the formula

$$EM_p = \left\{ \frac{1}{m} \sum_{i=1}^m |e_i/W_i|^p \right\}^{1/p}$$

in his local error analysis. The quantitative model that Limb uses for the human viewer includes some error filtering as well. Comparison of the simple RMSE as a measure of image quality with the best error measure predictions of the model shows that RMSE performs surprisingly well. This results, Limb explains, from the fact that in most distorted images, quality is determined mainly by the visibility of distortion in flat areas where it is more visible and consequently the effects of masking have little effect. For images where distortion is greater at edges, however, the RMSE is claimed to be less satisfactory.

The results of a subjective evaluation on twelve versions of a black and white image and the rank ordering obtained with three computational measures are presented by Hall [22]. He compares the performance of the measures NMSE, LMSE, and PMSE, which are defined for an $N \times N$ discrete image as

$$NMSE = \frac{\sum_{m=1}^N \sum_{n=1}^N [f(m,n) - \hat{f}(m,n)]^2}{\sum_{m=1}^N \sum_{n=1}^N [f(m,n)]^2}$$

$$LMSE = \frac{\sum_{m=2}^{N-1} \sum_{n=2}^{N-1} [G(m,n) - \hat{G}(m,n)]^2}{\sum_{m=2}^{N-1} \sum_{n=2}^{N-1} [G(m,n)]^2}$$

where $G(m,n) = f(m+1,n) + f(n-1,n) + f(m,n+1) + f(m,n-1) - 4f(m,n)$

$$PMSE = \frac{\sum_{m=1}^N \sum_{n=1}^N [z(m,n) - \hat{z}(m,n)]^2}{\sum_{m=1}^N \sum_{n=1}^N [z(m,n)]^2},$$

where $z(m,n)$ and $\hat{z}(m,n)$ are given by

$$z(m,n) = \ln[f(m,n)] \otimes h_{bp}(m,n)$$

and

$$\hat{z}(m,n) = \ln[\hat{f}(m,n)] \otimes h_{bp}(m,n)$$

The function $h_{bp}(m,n)$ is a rectangular coordinate form of the point spread function of the HVS. In his comparison, Hall finds that the correlation between PMSE and the subjective ranking (obtained by using bubble sort) of the data set is higher than that of NMSE and LMSE.

Nil [8] arrives at a quality measure in the 2-D discrete Fourier spatial frequency domain. This measure is expressed as

$$K^{-1} \sum_{i=1}^B W_i \sum_{u=0}^{M-1} \sum_{v=0}^{N-1} H^2(r) [F_i(u,v) - \hat{F}_i(u,v)]^2,$$

where B = number of subimage blocks in scene,

K = normalization factor such as total energy,

$H(r)$ = rotationally symmetric spatial frequency response of HVS, $r = \sqrt{u^2 + v^2}$,

F_i, \hat{F}_i = Fourier transform of unprocessed and processed subimage i , respectively,

M, N = number of Fourier coefficients + 1, in orthogonal u, v directions,

W_i = subimage i structure weighting factor, proportional to subimage's intensity level variance.

Using $H(r) = (0.2 + 0.45r)e^{-0.18r}$, he then constructs the function

$$|A(r)| H(r) = \begin{cases} 0.05r^{0.554}, & \text{for } r < 7 \\ e^{-9 [|\log_{10} r - \log_{10} 9|]^{2.3}} & \text{for } r \geq 7 \end{cases}$$

for dealing with image cosine transforms instead of image Fourier transforms. Finally, he argues that (i) combining the HVS model with the image cosine transform will result in better performance in image compression and image quality assessment applications, and (ii) performance in quality assessment should also be enhanced by inclusion of the subimage structure weighting.

Marmolin [9] addresses the question of using the mean squared error (MSE) measure as a quality criterion in image processing, and evaluates the predictive power of

$$E = \left[\frac{1}{n} \sum_{i=1}^n |D_i|^p \right]^{1/p},$$

$$D_i = a_i - g(x_i - y_i)$$

where g = some processing function that determines the visibility of the error, a_i = a weight related to the informative value of pixel i , and p = a factor that determines the relative importance of small and large errors, x_i = the gray level of pixel i in the original image, y_i = the gray level of pixel i in the processed image. He investigates the performance of different definitions for D_i , and compares them to that of the mean squared error

$$\text{MSE} = \left[\frac{1}{n} \sum_{i=1}^n (x_i - y_i)^2 \right]^{1/2}$$

The results obtained indicate that MSE is an unsatisfactory measure of perceived similarity, and that no measure is valid for each image set used.

Saghri, Cheatham, and Habibi [10] state that once an image $U(x,y)$ and its reproduction have been subjected to the HVS model, then the mean square error

$$d(U,U') = \frac{1}{N} \iint [U(x,y) - U'(x,y)] dx dy,$$

where N is the image area or the number of pixels, may be considered as a meaningful measure of image quality. Adopting the approach of Mannos and Sakrison, they use in their HVS model

$$f(u) = u^{0.33}$$

where u is the pixel intensity, and

$$A(f_r) = \left[0.2 + 0.81 \left(\frac{f_r}{5.55} \right) \right] \exp \left[- \left(\frac{f_r}{5.55} \right) \right],$$

where $f_r = (f_x^2 + f_y^2)^{1/2}$. The corrections (developed by Nill)

$$C(f_r) = \left(\frac{1}{4} + \frac{1}{2} \left\{ \log_e \left[\frac{2\Pi f_r}{\alpha} + \left(\frac{4\Pi^2 f_r^2}{\alpha} + 1 \right)^{1/2} \right] \right\}^2 \right)^2$$

to the HVS model of $A(f_r)$ is then added to give the DCT version

$$A_{DCT} = A(f_r)C(f_r).$$

As an alternative to the MSE, the authors propose the so-called information content (IC). The IC of an image for a given resolution is defined as the sum of the magnitudes of its DCT spectral components after they have been appropriately normalized based on HVS sensitivity models for that particular resolution. The plot of IC versus the resolution provides some insight into the quality of a given image. The preliminary results are reportedly promising, but much more experimentation is needed to adjust the numerous parameters of the system for highest achievable correlation with the subjective measure.

The work by Ngan, Leong, and Singh [16] describes an adaptive cosine transform coding scheme for color images. The cosine transform coefficients are weighted by the HVS function given by Nill to generate the coefficients in perceptual domain. To determine the parameters of the HVS filter

$$H(w) = (a+b\omega) \exp(-c\omega)$$

plots of SNR versus peak frequency are used. The SNR is defined by

$$SNR = -10 \log_{10} \left[\frac{1}{(512)^2} \sum_{j=0}^{511} \sum_{k=0}^{511} \frac{[f(j,k) - \hat{f}(j,k)]^2}{(255)^2} \right],$$

where $f(j,k)$ and $\hat{f}(j,k)$ are the original and reconstructed pixels, respectively. Their results show that the subjective quality of the reconstructed images at a bit rate of 0.4 bit/pixel or a compression ratio of 60:1 is very good.

Khafizov, Fisher, and Kiselyov [18] propose a new approach to simulate human visual perception in order to devise a tool for measuring distance between images. Defining the error matrix by

$$E = X - Y,$$

where X and Y are the two images to be compared, they renormalize each error in E with respect to other errors. Renormalization is the core of their method and it produces a new re-estimated error matrix E' . Once E' is obtained, they compute the L_1 -norm of E' as the distance between X and Y . In the case when there are only two errors e and z in E , the formula

$$e'(z) = \frac{3+a^s}{z(1+a^s)} (e+\tilde{z}), \text{ where } \tilde{z} = \begin{cases} z, & ez > 0 \\ 2e-z, & ez < 0 \end{cases}$$

where a = some positive constant and s = distance between e and z , is used for re-estimating the error e with respect to error z . The generalization to an arbitrary case is immediate. The experiments presented demonstrate the inconsistency of the conventional RMSE together with the success in simulating visual human perception.

Nil and Bouzas [17] present an objective, quantitative image quality measure based on the digital image power spectrum of normally acquired arbitrary scenes. Using polar coordinates ρ , Θ the image quality measure is derived from the normalized 2-D power spectrum $P(\rho, \Theta)$ weighted by the square of the modulation transfer function of the human visual system $A^2(T\rho)$, the directional scale of the input image $S(\Theta_1)$, and the modified Wiener noise filter $W(\rho)$:

$$IQM = \frac{1}{M^2} \sum_{\Theta=-180}^{180} \sum_{\rho=0.01}^{0.5} S(\Theta_1)W(\rho)A^2(T\rho)P(\rho, \Theta),$$

where M^2 = number of pixels. In its application, a previously constructed modulation transfer function [8] is used for the HVS. The authors point out that the power spectrum approach does not require use of designed quality assessment targets or reimaging the same scene for comparison purposes. Experimental verification indicates good correlation of this objective quality measure with visual quality assessments.

3. Conclusions

Traditionally, the most reliable way of measuring image quality has been the subjective evaluation by human observers. Because of the inherent difficulties associated with this approach, much attention has been focused on the development of quantitative techniques for quick and objective measurement. The image quality measure that has been commonly used in digital image compression is the mean square error (MSE) between the original image and the reconstructed image. It is now a well-known fact, however, that the MSE and its variants do not correlate reasonably well with subjective quality measures [4, 5, 7-10, 21]. A major portion of recent research is, therefore, directed towards incorporating human visual system (HVS) models into image quality measures. This is not a trivial task because the human visual system is too complex and an accurate model cannot presently be developed. Nevertheless, a number of experiments with simplified models indicates that the inclusion of a model for the HVS generally produces results that are in better correlation with the perceived image quality [4, 7, 8, 10-18, 22]. The trial models take into consideration various recognized characteristics of the HVS, and usually have both linear

and nonlinear parts. As we have a better understanding of the psychophysical phenomena concerning the human vision, we will be able to develop more accurate models which, in turn, will lead to results closer to the human response.

References

- [1] T.C. Bell, J.G. Cleary, and I.H. Witten, *Text Compression*, Prentice-Hall, Inc., USA, 1990.
- [2] J.A. Storer (Editor), *Image and Text Compression*, Kluwer Academic Publishers, USA, 1992.
- [3] M. Nelson, *The Data Compression Book*, M&T Publishing, Inc., 1992.
- [4] A.K. Jain, "Image Data Compression: A Review," *Proceedings of the IEEE*, Vol. 69, No. 3, pp. 349-389, March 1981.
- [5] A.K. Jain, *Fundamentals of Digital Image Processing*, Prentice-Hall, Inc., USA, 1989.
- [6] M.P. Ekstrom (Editor), *Digital Image Processing Techniques*, Academic Press, Inc., USA, 1984.
- [7] D.J. Granrath, "The Role of the Human Visual Models in Image Processing," *Proceedings of the IEEE*, Vol. 69, No. 5, pp. 552-561, May 1981.
- [8] N.B. Nill, "A Visual Model Weighted Cosine Transform for Image Compression and Quality Assessment," *IEEE Transactions on Communications*, Vol. COM-33, No. 6, pp. 551-557, June 1985.
- [9] H. Marmolin, "Subjective MSE Measures," *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. SMC-16, No. 3, pp. 486-489, May/June 1986.
- [10] J.A. Saghri, P.S. Cheatham, and A. Habibi, "Image Quality Measure Based on a Human Visual System Model," *Optical Engineering*, Vol. 28, No. 7, pp. 813-818, July 1989.
- [11] J.L. Mannas, "The Effects of a Visual Fidelity Criterion on the Encoding of Images," *IEEE Transactions on Information Theory*, Vol. IT-20, No. 4, pp. 525-536, July 1974.
- [12] J.O. Limb, "Distortion Criteria of the Human Viewer," *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. SMC-9, No. 12, pp. 778-793, December 1979.
- [13] D.J. Sakrison, "On the Role of the Observer and a Distortion Measure in Image Transmission," *IEEE Transactions on Communications*, Vol. COM-25, No. 11, pp. 1251-1267, November 1977.
- [14] A.N. Netravali and J.O. Limb, "Picture Coding: A Review," *Proceedings of the IEEE*, Vol. 68, No. 3, pp. 366-406, March 1980.
- [15] F.X.J. Lukas and Z.L. Budrikis, "Picture Quality Prediction Based on a Visual Model," *IEEE Transactions on Communications*, Vol. COM-30, No. 7, pp. 1679-1692, July 1982.

- [16] K.N. Ngan, K.S. Leong, and H. Singh, "Adaptive Cosine Transform Coding of Images in Perceptual Domain," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol 37, No. 11, pp. 1743-1750, December 1989.
- [17] N.B. Nill and B.H. Bouzas, "Objective Image Quality Measures Derived From Digital Image Power Spectra," *Optical Engineering*, Vol. 31, No. 4, pp. 813-825, April 1992.
- [18] F.T. Khafizor, P.S. Fisher, and O. Kiselyov, "A Note on Comparing Images," submitted to *IEEE Computing*, 1992.
- [19] W.K. Pratt, *Digital Image Processing*, John Wiley and Sons, Inc., USA, 1978.
- [20] N.B. Nill, "Scene Power Spectra: The Moment as an Image Quality Merit Factor," *Applied Optics*, Vol. 15, No. 11, pp. 2846-2854, November 1976.
- [21] D.R. Ahlgren, J. Crosbie, and D. Erigat, "Compression of Digitized Images for Transmission and Storage Applications," *Proceedings of SPIE*, Vol. 901, pp. 105-113, 1988.
- [22] C.F. Hall, "Subjective Evaluation of a Perceptual Quality Metric," *Proceedings of SPIE*, Vol. 310, pp. 200-204, 1981.
- [23] H.L. Snyder, "Image Quality: Measures and Visual Performance," *Flat-Panel Displays and CRTs*, L.E. Tannas, Jr., Ed., Van Nostrand Reinhold, New York, , pp. 70-90, 1985.
- [24] P.M. Farrelle, *Recursive Block Coding for Image Data Compression*, Springer-Verlag New York Inc., USA, 1990.
- [25] P.C. Cosman, C. Tseng, R.M. Gray, R.A. Olshen, L.E. Moses, H.C. Davidson, C.J. Bergin, and E.A. Riskin, "Tree-structured Vector Quantization of CT Chest Scans: Image Quality and Diagnostic Accuracy," Technical Report No. 157, Division of Biostatistics, Stanford University, Stanford, California, August 1992.

